View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by PhilSci Archive

Failure of psychophysical supervenience in many worlds

July 1, 2018

Abstract

Psychophysical supervenience requires that the mental properties of a system cannot change without the change of its physical properties. In this paper, I argue that the Everett interpretation of quantum mechanics or Everett's theory seems to violate the principle of psychophysical supervenience. In order to be consistent with our experience, the theory assumes psychophysical supervenience in each world, including our world. However, this permits the possibility that under certain unitary time evolution which does not lead to world branching, the wave function of each world changes and correspondingly the mental states of the observers in the world also change, while the wave function of the total worlds does not change, which violates the principle of psychophysical supervenience for all worlds. It seems that one must go beyond Everett's theory such as denying multiplicity in order

to avoid the failure of psychophysical supervenience.

Psychophysical supervenience is an important principle in the philosophy of mind. The standard definition of supervenience is that a set of properties A supervenes on another set B in case no two things can differ with respect to A-properties without also differing with respect to their B-properties (see McLaughlin and Bennett, 2014). By this definition, psychophysical supervenience requires that the mental properties of a system cannot change without the change of its physical properties. In particular, for a system with many observers, the principle requires that the mental properties of each observer cannot change without the change of the physical properties of the system. In this paper, I will argue that the Everett interpretation of quantum mechanics or Everett's theory seems to violate the principle of psychophysical supervenience.

Everett's theory assumes that the wave function of a physical system is a complete description of the system, and the wave function always evolves in accord with the linear Schrödinger equation. In order to solve the measurement problem, the theory further assumes that after a measurement with many possible results there appear many equally real worlds, in each of which there is an observer who is consciously aware of a definite result (Everett, 1957; DeWitt and Graham, 1973; Wallace, 2012). In the following, I will analyze whether Everett's theory is consistent with the principle of

psychophysical supervenience.

Consider a simple measurement situation, in which an observer M interacts with the measured system S. When the state of S is $|0\rangle_S$[1], the state of M does not change after the interaction:

$$|0\rangle_S |ready\rangle_M \rightarrow |0\rangle_S |ready\rangle_M . \tag{1}$$

When the state of S is $|1\rangle_S$, the state of M changes and she obtains a measurement result:

$$|1\rangle_S |ready\rangle_M \rightarrow |1\rangle_S |1\rangle_M . \tag{2}$$

The interaction can be represented by a unitary time evolution operator, U. Then the above two processes can be formulated as follows:

$$U |0\rangle_S |ready\rangle_M = |0\rangle_S |ready\rangle_M . \tag{3}$$

$$U |1\rangle_S |ready\rangle_M = |1\rangle_S |1\rangle_M . \tag{4}$$

According to Everett's theory, there is no world branching, and there is still one observer, namely the original observer, after these evolution. In the first case, after the interaction, the physical state of the observer, including

[1]I will use the elegant Dirac notation throughout this paper.

her memory, does not change, and her mental state does not change either. In the second case, after the interaction, the physical state of the observer, including her memory, changes from $|ready\rangle_M$ to $|1\rangle_M$, and correspondingly her mental state changes from the ready state to a result state; she obtains the result 1. Moreover, the observer is also consciously aware of the change of her mental state. This is a valid measurement. Obviously, the principle of psychophysical supervenience is satisfied in these processes.

Now suppose the observer M interacts with the system S being in a superposed state, $|0\rangle_S + |1\rangle_S$. For simplicity I omit the nomalization factor $1/\sqrt{2}$. By the linear Schrödinger equation, the physical state of the composite system after the interaction will evolve into the following superposition:

$$|0\rangle_S |ready\rangle_M + |1\rangle_S |1\rangle_M . \tag{5}$$

That is:

$$U(|0\rangle_S + |1\rangle_S) |ready\rangle_M = |0\rangle_S |ready\rangle_M + |1\rangle_S |1\rangle_M . \tag{6}$$

According to Everett's theory, there is world branching after this interaction, and the post-measurement state corresponds to two worlds, in each of which there is an observer who has a definite perception, either being in the ready state or obtaining result 1.[2]

[2]Here I omit the environment terms in the evolution, which, in a more complete form, should be $U(|0\rangle_S + |1\rangle_S) |ready\rangle_M |ready\rangle_E = |0\rangle_S |ready\rangle_M |ready\rangle_E + |1\rangle_S |1\rangle_M |1\rangle_E$. Besides, it may be worth noting that in Wallace's (2012) formulation of Everett's theory

There are in general three ways of understanding the notion of multiplicity in Everett's theory: (1) measurements lead to multiple worlds at the fundamental level (DeWitt and Graham, 1973), (2) measurements lead to multiple worlds only at the non-fundamental "emergent" level (Wallace, 2012), and (3) measurements only lead to multiple minds (Zeh, 1981).[3] In either case, for the above post-measurement state (6), the mental state of each observer is not determined uniquely by the whole wave function, but determined only by the corresponding branch of the wave function.[4] This means that the principle of psychophysical supervenience is satisfied in each world. The question is: is the principle of psychophysical supervenience satisfied in the whole worlds?

In order to answer this question, let's analyze possible evolution of the above post-measurement state or the corresponding worlds. First, consider a unitary time evolution operator, U_A, which does not lead to world branching and changes $|0\rangle_S |ready\rangle_M$ to $|1\rangle_S |1\rangle_M$ and $|1\rangle_S |1\rangle_M$ to $|0\rangle_S |A_0\rangle_M$:

$$U_A |0\rangle_S |ready\rangle_M = |1\rangle_S |1\rangle_M , \tag{7}$$

the number of the emergent observers after a measurement is not definite due to the imperfectness of decoherence. My following analysis also applies to this formulation.

[3]It is worth noting that Albert and Loewer's (1988) many-minds theory does not assume the usual notion of multiplicity as listed above. It assumes the existence of infinitely many minds even for a post-measurement product state, and it already entails dualism and violates the principle of psychophysical supervenience. I will not discuss this theory in this paper.

[4]If the mental state of each observer is not determined by the corresponding branch of the post-measurement superposition, then the predictions of the theory will be not consistent with the predictions of quantum mechanics and experience for some unitary time evolution of the superposition.

$$U_A \left|1\right\rangle_S \left|1\right\rangle_M = \left|0\right\rangle_S \left|A_0\right\rangle_M , \tag{8}$$

where $\left|A_0\right\rangle_M$ is a definite mental state of M. The principle of psychophysical supervenience is satisfied for the evolution. In particular, the first evolution of the observer is exactly the same as the evolution of the observer in (4); the physical state of the observer changes from $\left|ready\right\rangle_M$ to $\left|1\right\rangle_M$, and correspondingly her mental state changes from the ready state to a result state. Moreover, the observer is also consciously aware of the change of her mental state.

Then the unitary time evolution of the above post-measurement state is

$$U_A(\left|0\right\rangle_S \left|ready\right\rangle_M + \left|1\right\rangle_S \left|1\right\rangle_M) = \left|1\right\rangle_S \left|1\right\rangle_M + \left|0\right\rangle_S \left|A_0\right\rangle_M . \tag{9}$$

By the linearity of the dynamics, the evolution of the two worlds are the same as the above two forms of evolution, and the principle of psychophysical supervenience is satisfied in each world. For example, when the physical state of the observer in the first world changes from $\left|ready\right\rangle_M$ to $\left|1\right\rangle_M$, her mental state changes from the ready state to a result state. Moreover, she is consciously aware of the change of her mental state.

Now consider one of these unitary time evolution operators, U_N, for which $\left|A_0\right\rangle_M = \left|ready\right\rangle_M$. In other words, U_N changes $\left|0\right\rangle_S \left|ready\right\rangle_M$ to $\left|1\right\rangle_S \left|1\right\rangle_M$ and $\left|1\right\rangle_S \left|1\right\rangle_M$ to $\left|0\right\rangle_S \left|ready\right\rangle_M$. It is similar to the NOT gate for

a single q-bit, and is permitted by the Schrödinger equation in principle. Then the unitary time evolution of the above post-measurement state is

$$U_N(|0\rangle_S |ready\rangle_M + |1\rangle_S |1\rangle_M) = |1\rangle_S |1\rangle_M + |0\rangle_S |ready\rangle_M . \qquad (10)$$

Again, there is no world branching, and the principle of psychophysical supervenience is satisfied in each world. When the physical state of the observer in the first world changes from $|ready\rangle_M$ to $|1\rangle_M$, her mental state changes from the ready state to a result state; she obtains the result 1. Moreover, she is consciously aware of the change of her mental state. Similarly, when the physical state of the observer in the second world changes from $|1\rangle_M$ to $|ready\rangle_M$, her mental state also changes from the result state to the ready state correspondingly; she "loses" the result 1 (and relevant memory). On the other hand, after the unitary time evolution the whole superposition does not change.

Therefore, Everett's theory predicts that after the above unitary time evolution, the physical state of the composite system, which is completely represented by the wave function of the system, does not change, while the mental states of the two involved observers both change after the evolution. This means that the principle of psychophysical supervenience, which requires that the mental properties of a system cannot change without the

change of its physical properties, is violated for the evolution of the whole worlds.

There are two possible ways to avoid the violation of psychophysical supervenience. The first way is to deny that after the evolution the physical state of the composite system has not changed. This requires that the wave function of a system is not a complete description of the physical state of the system, and additional variables are needed to introduce to describe the complete physical state. However, this requirement is not consistent with Everett's theory. Moreover, it is worth noting that in order to save psychophysical supervenience, it is also required that the additional variables should be changed by the unitary time evolution of the wave function, and the mental state of an observer should also supervene on the additional variables; otherwise the introduction of these variables cannot save psychophysical supervenience in the above example.

The second way to avoid the violation of psychophysical supervenience in the above example is to deny that after the evolution the total mental properties of the composite system have changed. However, this is inconsistent with the requirement of Everett's theory that the mental state of the observer and its evolution in each world is determined by the corresponding branch of the post-measurement superposition and its evolution (when there is no world branching in this world). As noted before, this requirement is necessiated by the linearity of the dynamics and the consistency of the

theory with our experience in our world. By this requirement, when the physical state of the observer in a world changes from $|ready\rangle_M$ to $|1\rangle_M$, her mental state also changes from the ready state to the corresponding result state. Moreover, the observer is consciously aware of the change of her mental state. Since the mental state of the observer in each world changes, the mental properties of the composite system also change.

In order to see this point more clearly, one can compare the above unitary time evolution operator U_N and the identity operator I:

$$I(|0\rangle_S |ready\rangle_M + |1\rangle_S |1\rangle_M) = |0\rangle_S |ready\rangle_M + |1\rangle_S |1\rangle_M . \qquad (11)$$

Under the identity time evolution I, the mental state of the observer in each world does not change. While under the unitary time evolution U_N, the mental state of the observer in each world changes, such as from the ready state to a result state, and the observer is also consciously aware of the change of her mental state; she obtains a result. On the other hand, if there are no many worlds and only one world, then after the above evolution U_N the total mental properties of the composite system do not change either, since after the evolution there remains an observer with the same mental content, which contains a null result (ready state) and a result 1. This is the situation in collapse theories (see Gao, 2017).

Finally, it may be worth noting that since there is no world branching during the above evolution, there is no additional complexity about the

definition of the identity of observers, and the analysis of personal identity is the same as that in the classical world (see, e.g. Olson, 2017). If one denies that the observer in each of the two worlds keeps her identity after the above evolution, then one need to either deny the linearity of the dynamics or deny the observer keeps her identity even when she makes a measurement in a single world (see (4) or (7)). In fact, no matter how to define personal identity, if only the mental state of an observer changes from the ready state to a result state, the observer can be consciously aware of the change of the mental state. While if the mental state of an observer does not change, the observer can certainly not be consciously aware of the change. This difference is the main basis of the above analysis.

To sum up, I have argued that Everett's theory seems to violate the principle of psychophysical supervenience. In order to be consistent with our experience, the theory assumes psychophysical supervenience in each world, including our world. For example, minds are emergent in each world according to Wallace's (2012) formulation. This permits the possibility that under certain unitary time evolution each world branch changes and correspondingly the mental states of the observers in the world also changes, while the wave function of the total worlds does not change. This possibility leads to the violation of psychophysical supervenience for all worlds, since the wave function of a system is a complete description of the physical state of the system in Everett's theory. It seems that one must go beyond

Everett's theory such as denying multiplicity in order to avoid the failure of psychophysical supervenience.

References

[1] Albert, D. Z. and B. Loewer. (1988). Interpreting the Many Worlds Interpretation, Synthese, 77, 195-213.

[2] DeWitt, B. S. and N. Graham (eds.). (1973). The Many-Worlds Interpretation of Quantum Mechanics. Princeton: Princeton University Press.

[3] Everett, H. (1957). 'Relative state' formulation of quantum mechanics. Rev. Mod. Phys. 29, 454-462.

[4] Gao, S. (2017). The Meaning of the Wave Function: In Search of the Ontology of Quantum Mechanics. Cambridge: Cambridge University Press.

[5] McLaughlin, B. and Bennett, K. (2014). Supervenience, The Stanford Encyclopedia of Philosophy (Spring 2014 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/spr2014/entries/supervenience/.

[6] Olson, E. T. (2017). Personal Identity, The Stanford Encyclopedia of Philosophy (Summer 2017 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/sum2017/entries/identity-personal/.

[7] Wallace, D. (2012). The Emergent Multiverse: Quantum Theory according to the Everett Interpretation. Oxford: Oxford University Press.

[8] Zeh, H. D. (1981). The Problem of Conscious Observation in Quantum Mechanical Description, Epistemological Letters of the Ferdinand-Gonseth Association in Biel (Switzerland), 63. Also Published in Foundations of Physics Letters. 13 (2000) 221-233.

2.0